From multilinear SVD to multilinear UTV decomposition*

Michiel Vandecappelle[a,b,1,*], Lieven De Lathauwer[a,b]

[a] *KU Leuven, Dept. of Electrical Engineering ESAT/STADIUS, Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven, Belgium*
[b] *KU Leuven – Kulak, Group Science, Engineering and Technology, E. Sabbelaan 53, B-8500 Kortrijk, Belgium*

Abstract

Across a range of applications, low multilinear rank approximation (LMLRA) is used to compress large tensors into a more compact form, while preserving most of their information. A specific instance of LMLRA is the multilinear singular value decomposition (MLSVD), which can be used for multilinear principal component analysis (MLPCA). MLSVDs are obtained by computing SVDs of all tensor unfoldings, but, in practical applications, it is often not necessary to compute full SVDs. In this article, we therefore propose a new decomposition, called the truncated multilinear UTV decomposition (TMLUTVD). This is a tensor decomposition that is also multilinear rank-revealing, yet less expensive to compute than a truncated MLSVD (TMLSVD); it can even be computed in a finite number of steps. We present its properties in an algorithm-independent manner. In particular, we derive bounds on the accuracy in function of the truncation level. Experiments illustrate the good performance in practice.

Keywords: Tensor, Multilinear SVD, Subspace analysis

1. Introduction

Tensors have become increasingly popular tools for applications in signal processing [5] and machine learning [21]. Tensors are higher-order extensions of vectors and matrices which allow higher-order data to be stored while preserving their inherent higher-order structure. Many different tensor decompositions have been proposed to facilitate analysis and compression of tensor data. The most common decompositions are the Canonical Polyadic Decomposition (CPD) [15]

*Research supported by: (1) Flemish Government: This work was supported by the Fonds de la Recherche Scientifique–FNRS and the Fonds Wetenschappelijk Onderzoek–Vlaanderen under EOS project no 30468160 (SeLMA) and by Artificiële intelligentie (AI) Vlaanderen (3E190661); (2) KU Leuven Internal Funds C16/15/059 and ID-N project no 3E190402; (3) Leuven Institute for Artificial Intelligence (Leuven.ai).

*Corresponding author

[1] Michiel Vandecappelle was partially supported by an SB Grant (1S66617N) from the Research Foundation—Flanders (FWO).

and the Low Multilinear Rank Approximation (LMLRA) [23]; for the latter, see Figure 1. The Multilinear Singular Value Decomposition (MLSVD) [6] is a special case of LMLRA that can be obtained algebraically: for each mode, an SVD is computed of the mode-n unfolding of the tensor and, when necessary, a core tensor that links the different modes is constructed through a multilinear projection. The excellent numerical properties of the SVD are inherited by the MLSVD and the MLSVD can be truncated as a natural higher-order extension of the truncated SVD. This truncated MLSVD (TMLSVD) offers an LMLRA that is much cheaper to compute than the optimal LMLRA for the given rank, while its approximation error is often close to minimal [6, 27]. Because of this, the TMLSVD is widely used in tensor applications.

When considering a noisy low-rank tensor, one is often only interested in extracting the principal subspace of the tensor in every mode: all multilinear singular vectors in the principal subspaces are retained, while those corresponding to the noise are trimmed. If extraction of the principal subspaces is the only requirement, a cheaper rank-revealing decomposition can be used instead of the SVD. For example, a rank-revealing UTV (RRUTV) decomposition, wherein the middle matrix is triangular instead of diagonal for the SVD, provides a basis for the principal subspace that is less expensive to compute than the basis obtained by the SVD. The bases for the principal subspaces in every mode can then be used as factor matrices for an LMLRA. Afterwards, the core tensor can be computed, as for the MLSVD, leading to a truncated multilinear UTV decomposition (TMLUTVD). Because of its lower computational cost compared to the TMLSVD, the TMLUTVD is particularly interesting for tensor tracking applications, where computation time is scarce and the principal mode-n subspaces are often the prime interest, see [4, 25].

The article is structured as follows. In Section 2, the TMLUTVD is introduced. Next, bounds for the approximation error are derived and proved in Section 3. Finally, Section 4 demonstrates that the TMLUTVD offers an LMLRA with similar accuracy as the TMLSVD in practice. The decomposition is laid out for real third-order tensors. Evidently, the TMLUTVD can also be computed for complex tensors by adding complex conjugations in the appropriate places. Extensions to higher orders are straightforward as well.

1.1. Notation and terminology

Scalars, vectors and matrices are denoted by lowercase (a), bold lowercase ($\mathbf{a}$) and bold uppercase letters ($\mathbf{A}$), respectively, while tensors are written in calligraphic script ($\mathcal{A}$). The identity matrix and the zero matrix are denoted by $\mathbf{I}$ and $\mathbf{0}$, respectively. The Kronecker product of matrices is written as $\otimes$. When using subscripts to specify certain entries of a tensor $\mathcal{T}$, a

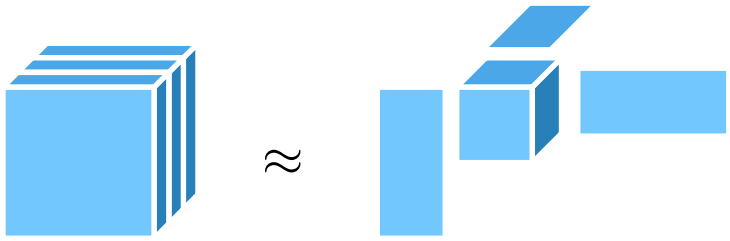

Figure 1: LMLRA of a third-order tensor. The tensor is decomposed into three factor matrices, linked by a core tensor.

colon refers to all entries of a specific mode, e.g. $\mathbf{t}_{1:2}$ is a vector holding all entries of the third-order tensor $\mathcal{T}$ which have 1 and 2 as their first and third index, respectively. More general, a mode-n fiber of a tensor is a vector obtained by fixing all indices but the nth. The mode-n unfoldings $\mathbf{T}_{(n)}$ of a third-order tensor $\mathcal{T}$ are constructed by stacking all its mode-n fibers as columns of a matrix, with indices of lower modes running faster. For example: in the mode-2 unfolding $\mathbf{T}_{(2)}$, the column ordering of the mode-2 fibers runs along mode 3 and mode 1, in that order, thus $t_{i_1 i_2 i_3}$ is mapped to $\left(\mathbf{T}_{(2)}\right)_{i_2,q}$, with $q = i_1 + (i_3 - 1)I_1$. The mode-n tensor-matrix product of a tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and a matrix $\mathbf{X} \in \mathbb{R}^{J \times I_n}$ is written as $\mathcal{T} \cdot_n \mathbf{X}$. For $n = 1$, 2 and 3, $\mathcal{T} \cdot_n \mathbf{X} \in \mathbb{R}^{J \times I_2 \times I_3}$, $\mathbb{R}^{I_1 \times J \times I_3}$ and $\mathbb{R}^{I_1 \times I_2 \times J}$, respectively. In unfolded form, computing the mode-n product is the same as computing $\mathbf{X}\mathbf{T}_{(n)}$. Products of the form $\mathbf{X}(\mathbf{Z} \otimes \mathbf{Y})$ are never explicitly formed, as they can be computed as matricized tensor times Kronecker products (MTKRONPRODs), for which specialized routines exist [26]. It suffices to store the individual matrices $\mathbf{Y}$ and $\mathbf{Z}$.

An LMLRA of a tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is defined as $\mathcal{T} \approx \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$, where $\mathcal{S} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$, $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times R_1}$, $\mathbf{U}_2 \in \mathbb{R}^{I_2 \times R_2}$, $\mathbf{U}_3 \in \mathbb{R}^{I_3 \times R_3}$. The vector (R_1, R_2, R_3), with $R_n \leq I_n$, holds the number of (linearly independent) columns in the factor matrices $\mathbf{U}_n$. If the approximation is exact and the $R^{(i)}$ are minimal, then this vector is called the multilinear rank.

2. Multilinear UTV decomposition

The MLSVD [6,23] decomposes an Nth-order tensor into a set of N orthogonal factor matrices, which are obtained through N SVDs. These factor matrices are linked by a core tensor $\mathcal{S}$, found by forming $\mathcal{S} = \mathcal{T} \cdot_1 \mathbf{U}_1^{\mathrm{T}} \cdot_2 \mathbf{U}_2^{\mathrm{T}} \cdot_3 \mathbf{U}_3^{\mathrm{T}}$. The MLSVD is a highly structured decomposition that reveals a lot of information about the tensor. For each mode, the (vectorizations of the) mode-n slices of the core tensor are orthogonal to each other and ordered according to their norms, also called mode-n multilinear singular values. Due to the latter property, the MLSVD

can be truncated in any mode by dropping the factor matrix columns corresponding to the smallest multilinear singular values for that mode. Regardless of the number of truncated columns, the TMLSVD offers an LMLRA that is often close to optimal for the given multilinear rank [6]. The truncation can either be performed in parallel, i.e., for each mode the SVD is computed and truncated independently, or sequentially, i.e., the truncations are interleaved with the computation of the SVDs for the different modes, which decreases the tensor dimensions in every step and makes the decompositions for the following modes less costly. These approaches result in the TMLSVD [6] and the sequentially truncated MLSVD (STMLSVD) [27], respectively. STMLSVD computation is generally more efficient, but not necessarily more accurate [27].

When time is limited, computing an MLSVD can be prohibitively expensive. Also, it is often unnecessary to compute all information that the MLSVD offers. Alternative decompositions, such as the higher-order LQ (HOLQ) decomposition [9], a higher-order generalization of the LQ factorization wherein the factor matrices are lower triangular, have been proposed to extract only a relevant part of the multilinear information from the tensor. In the same spirit, we show in this article that instead of finding all multilinear singular values and allowing truncation to arbitrary multilinear rank, it can be sufficient to split each mode-n subspace of dimension I_n into its principal subspace, say of dimension R_n, and the orthogonal complement of dimension $(I_n - R_n)$ [7]. Ideally, the mode-n principal subspace corresponds to the underlying signal, while vectors in the orthogonal complement correspond to noise. We propose a new tensor decomposition that can be computed with the use of such cheaper alternatives for the SVD. In each mode, the principal subspace is extracted using a matrix decomposition that is cheaper to compute than the SVD when explicitly computing all multilinear singular values is not needed. In fact, one can even use a decomposition that can be computed in a finite number of steps instead of the SVD whose computation is intrinsically an iterative process. One can use a UTV-type decomposition such as the rank-revealing QR [2, 10] (RRQR) decomposition, the more general RRUTV [7, 22] decomposition or the hyperbolic [24] (HURV) or signed [28] (SURV) URV decomposition. These decompositions factorize a matrix $\mathbf{X}$ into a product $\mathbf{UTV}^{\mathrm{T}}$, with $\mathbf{U}$ and $\mathbf{V}$ orthogonal ($\mathbf{V}$ is J-orthogonal for the HURV and SURV decomposition) and $\mathbf{T}$ upper (URV decomposition) or lower triangular (ULV decomposition). Our presentation will be in terms of the rank-revealing URV decomposition (RRURVD), but other UTV decompositions can be used as well. Like the bottom right part of the diagonal matrix of the SVD, the bottom right corner of the $\mathbf{R}$-matrix of an RRURVD corresponds to the non-principal subspace and can be discarded to obtain a low-rank approximation of the matrix.

By combining the RRURVDs for the different tensor unfoldings of $\mathcal{T}$, one obtains an MLURVD.

Definition 1 (Multilinear URV decomposition). Let $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and let the orthogonal $\mathbf{U}_n$ be taken from RRURVDs of $\mathbf{T}_{(n)}$, for $n = 1, 2, 3$. Finally, let the core tensor $\mathcal{S} = \mathcal{T} \cdot_1 \mathbf{U}_1^{\mathrm{T}} \cdot_2 \mathbf{U}_2^{\mathrm{T}} \cdot_3 \mathbf{U}_3^{\mathrm{T}}$. Then

$$\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$$

is a multilinear URV decomposition (MLURVD) of $\mathcal{T}$.

Each factor matrix of an MLSVD can be split into parts corresponding to the principal mode-n subspace and to its orthogonal complement, respectively. The factor matrices of the MLURVD satisfy this property as well.

Property 1. *For the factor matrices $\mathbf{U}_n$ of an MLURVD $\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$, the first R_i columns correspond to the principal mode-n subspace of $\mathcal{T}$, while the last $(I_n - R_n)$ columns correspond to the orthogonal complement.*

For the core tensor $\mathcal{S}$ of an MLSVD, one has (outlined for mode-1, but analogous for the other modes):

$$\begin{aligned} &\mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} = \mathbf{T}_{(1)} = \mathbf{U}_1 \mathbf{S}_{(1)} (\mathbf{U}_3 \otimes \mathbf{U}_2)^{\mathrm{T}} \\ &\Leftrightarrow \mathbf{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} (\mathbf{U}_3 \otimes \mathbf{U}_2) = \mathbf{S}_{(1)}, \end{aligned}$$

wherein $\mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}}$ denotes the SVD of $\mathbf{T}_{(1)}$. Consequently, each row of $\mathbf{S}_{(1)}$, representing an $(I_2 \times I_3)$-slice $\mathbf{S}_{i::}$, is a scaled version of the corresponding row of $\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2)$, implying that all slices $\mathbf{S}_{i::}$ are orthogonal w.r.t. the standard matrix inner product. For the MLURVD, the subspaces are nested, so one has the following, more general, property.

Property 2. *For an MLURVD $\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$, one has $\mathbf{R}_1 \mathbf{V}_1^T (\mathbf{U}_3 \otimes \mathbf{U}_2) = \mathbf{S}_{(1)}$. The upper triangularity of $\mathbf{R}_1$ then implies that the last row of $\mathbf{S}_{(1)}$ is a scaled version of the last row of $\mathbf{V}_1^T (\mathbf{U}_3 \otimes \mathbf{U}_2)$, the second-to-last row of $\mathbf{S}_{(1)}$ is generated by the last two orthonormal rows of $\mathbf{V}_1^T (\mathbf{U}_3 \otimes \mathbf{U}_2)$, i.e., one orthogonal direction is added, and so on.*

Similar to the approach for the TMLSVD, the MLURVD allows for a truncation to a low multilinear rank-$(R_1, R_2, \ldots, R_N)$ approximation by considering the mode-n principal subspaces, see Figure 2. One retains the R_n columns of the $\mathbf{U}_n$-matrices corresponding to the principal mode-n subspaces and the associated parts of the core $\mathcal{S}$, leading to an approximation of $\mathcal{T}$.

Definition 2 (Truncated MLURVD). From an MLURVD

$$\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$$

of a tensor $\mathcal{T}$, a truncated MLURVD (TMLURVD) is obtained by dropping all but the first R_n columns of each $\mathbf{U}_n$-matrix, along with the corresponding parts of $\mathcal{S}$.

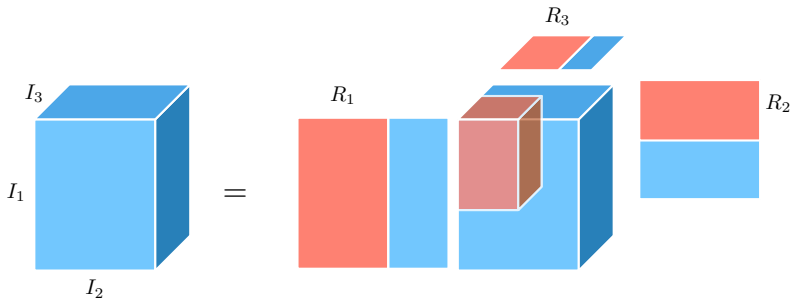

Figure 2: Both the MLURVD and the MLSVD can be truncated to obtain an LMLRA. This is achieved by retaining only the factor matrix columns and core tensor entries corresponding to the principal mode-n subspaces, marked in red. The relative importance of the columns is given by the mode-n singular values and the Frobenius norms of the rows of the upper-triangular $\mathbf{R}_n$-matrices for the MLSVD and the MLURVD, respectively.

A TMLURVD is less expensive to compute than a TMLSVD, but this comes at a cost. While the TMLSVD offers the minimal per-mode truncation error in the least squares sense for truncation to any multilinear rank, the MLURVD truncation error can be bounded, as explained in Section 3. MLURVD truncation can be either be done in parallel or sequentially, as UTV counterparts of the TMLSVD and the STMLSVD, respectively; see above.

3. Error bounds for the TMLURVD

The truncation of the MLURVD is performed by discarding all information that contributes less than a specified threshold to the tensor. We can bound this truncation error by looking at the norms of the parts of the triangular matrices that are discarded. Below, we derive error bounds for the TMLURVD and STMLURVD in both the Frobenius and spectral norms and explain how to choose appropriate theshold values.

3.1. TMLURVD

The $\mathbf{U}_n$-matrices of an MLURVD are orthogonal, so multiplication with them does not change the norm of a matrix. Thus, we only need to look at the core to derive accuracy bounds [6]. In Figure 3, we show that the triangle inequality can be used to bound the squared Frobenius norm of the truncated part of the core. An upper bound is given by the sum of the squared Frobenius norms of the discarded parts due to the truncations in the different tensor modes [6]. Writing ϵ_{F}^2 for the squared approximation error $\|\mathcal{T} - \hat{\mathcal{T}}\|_{\mathrm{F}}^2$, we have:

$$\epsilon_{\mathrm{F}}^2 \leq \|\mathbf{S}'_{(1)}\|_{\mathrm{F}}^2 + \|\mathbf{S}'_{(2)}\|_{\mathrm{F}}^2 + \|\mathbf{S}'_{(3)}\|_{\mathrm{F}}^2 \tag{1}$$

wherein $\mathbf{S}'_{(n)}$ denotes the matrix $\mathbf{S}_{(n)}$ without its first R_n rows. Property 2 tells us that $\mathbf{S}_{(1)} = \mathbf{R}_1\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2)$; the expressions for the mode-2 and mode-3 matrix representations

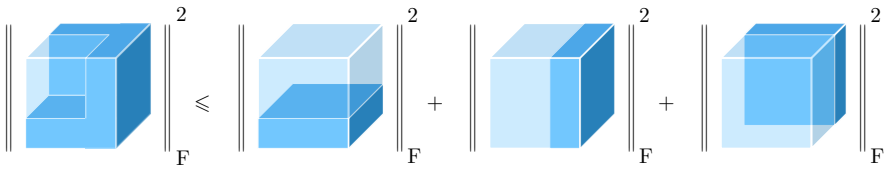

Figure 3: For the TMLURVD, the squared Frobenius norm of the truncated part of the core tensor is upper-bounded by the sum of the squared errors introduced by the truncations in the different modes.

are analogous. We can split the upper triangular $\mathbf{R}_n$-matrix into

$$\mathbf{R}_n = \begin{bmatrix} \mathbf{E}_n & \mathbf{F}_n \\ \mathbf{0} & \mathbf{G}_n \end{bmatrix} \tag{2}$$

with $\mathbf{E}_n \in \mathbb{R}^{R_n \times R_n}$ and $\mathbf{E}_n$ and $\mathbf{G}_n$ upper triangular. Dropping the first R_1 rows, we obtain

$$\mathbf{S'}_{(1)} = [\mathbf{0} \;\; \mathbf{G}_1]\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2),$$

in which the matrix $\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2) \in \mathbb{R}^{I_2 I_3 \times I_2 I_3}$ is row-wise orthonormal. By substituting the expressions for all $\mathbf{S'}_{(n)}$ matrices in (1), one obtains an upper bound for ϵ_{F}^2:

$$\begin{aligned} \epsilon_{\mathrm{F}}^2 &\leq \|[\mathbf{0} \;\; \mathbf{G}_1]\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2)\|_{\mathrm{F}}^2 + \|[\mathbf{0} \;\; \mathbf{G}_2]\mathbf{V}_2^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_1)\|_{\mathrm{F}}^2 \\ &\quad + \|[\mathbf{0} \;\; \mathbf{G}_3]\mathbf{V}_3^{\mathrm{T}}(\mathbf{U}_2 \otimes \mathbf{U}_1)\|_{\mathrm{F}}^2 \\ &= \|\mathbf{G}_1\|_{\mathrm{F}}^2 + \|\mathbf{G}_2\|_{\mathrm{F}}^2 + \|\mathbf{G}_3\|_{\mathrm{F}}^2. \end{aligned}$$

This bound clearly generalizes the bound for the TMLSVD in [6]: the $\mathbf{R}_n$ matrices are diagonal for the MLSVD and hold the mode-n multilinear singular values $\sigma_{n_{i_n}}$, yielding the bound

$$\epsilon_{\mathrm{F}}^2 \leq \sum_{i_1=R_1+1}^{I_1} \sigma_{1_{i_1}}^2 + \sum_{i_2=R_2+1}^{I_2} \sigma_{2_{i_2}}^2 + \sum_{i_3=R_3+1}^{I_3} \sigma_{3_{i_3}}^2.$$

The above triangle inequality also holds in the tensor spectral norm [8], defined as

$$\|\mathcal{A}\|_2 = \max_{\|\mathbf{x}\|=\|\mathbf{y}\|=\|\mathbf{z}\|=1} \mathcal{A} \cdot_1 \mathbf{x} \cdot_2 \mathbf{y} \cdot_3 \mathbf{z}.$$

Together with the fact that the spectral norm of a tensor is bounded by the spectral norms of its unfoldings [14], this leads to a bound on the error $\epsilon_2 = \|\mathcal{T} - \hat{\mathcal{T}}\|_2$:

$$\begin{aligned} \epsilon_2 &\leq \|\mathbf{S'}_{(1)}\|_2 + \|\mathbf{S'}_{(2)}\|_2 + \|\mathbf{S'}_{(3)}\|_2 \\ &\leq \|[\mathbf{0} \;\; \mathbf{G}_1]\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2)\|_2 + \|[\mathbf{0} \;\; \mathbf{G}_2]\mathbf{V}_2^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_1)\|_2 \\ &\quad + \|[\mathbf{0} \;\; \mathbf{G}_3]\mathbf{V}_3^{\mathrm{T}}(\mathbf{U}_2 \otimes \mathbf{U}_1)\|_2 \\ &= \|\mathbf{G}_1\|_2 + \|\mathbf{G}_2\|_2 + \|\mathbf{G}_3\|_2. \end{aligned}$$

3.2. STMLURVD

For the STMLURVD, the RRURVDs are different from those of the TMLURVD due to the incremental truncations. We write $\breve{\epsilon}_{\mathrm{F}}^2$, $\breve{\epsilon}_2$, $\breve{\mathbf{R}}_n$ and $\breve{\mathbf{G}}_n$ for the STMLURVD alternatives of ϵ_{F}^2, ϵ_2, $\mathbf{R}_n$ and $\mathbf{G}_n$, respectively.

Following a similar reasoning as for the TMLURVD, $\breve{\epsilon}_{\mathrm{F}}^2$ is equal to the squared sum of the errors introduced by the truncations in the different modes [27], i.e, in the STMLURVD analogue of Figure 3, the truncated parts of the right hand side do not overlap. Using (2) and Property 2, one finds

$$\begin{aligned}\breve{\epsilon}_{\mathrm{F}}^2 &= \|[\mathbf{0}\ \ \breve{\mathbf{G}}_1]\mathbf{V}_1^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_2)\|_{\mathrm{F}}^2 + \|[\mathbf{0}\ \ \breve{\mathbf{G}}_2]\mathbf{V}_2^{\mathrm{T}}(\mathbf{U}_3 \otimes \mathbf{U}_1)\|_{\mathrm{F}}^2 \\ &\quad + \|[\mathbf{0}\ \ \breve{\mathbf{G}}_3]\mathbf{V}_3^{\mathrm{T}}(\mathbf{U}_2 \otimes \mathbf{U}_1)\|_{\mathrm{F}}^2 \\ &= \|\breve{\mathbf{G}}_1\|_{\mathrm{F}}^2 + \|\breve{\mathbf{G}}_2\|_{\mathrm{F}}^2 + \|\breve{\mathbf{G}}_3\|_{\mathrm{F}}^2.\end{aligned}$$

For the spectral norm, we explicitly write out the error of the STMLURVD and use the triangle inequality to bound $\breve{\epsilon}_2$:

$$\begin{aligned}\breve{\epsilon}_2 &= \|\mathcal{T} - \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3\|_2 \\ &= \|\mathcal{T} - \mathcal{T} \cdot_1 \mathbf{W}_1 \cdot_2 \mathbf{W}_2 \cdot_3 \mathbf{W}_3\|_2 \\ &\leq \|\mathcal{T} - \mathcal{T} \cdot_1 \mathbf{W}_1\|_2 + \|\mathcal{T} \cdot_1 \mathbf{W}_1 - \mathcal{T} \cdot_1 \mathbf{W}_1 \cdot_2 \mathbf{W}_2\|_2 \\ &\quad + \|\mathcal{T} \cdot_1 \mathbf{W}_1 \cdot_2 \mathbf{W}_2 - \mathcal{T} \cdot_1 \mathbf{W}_1 \cdot_2 \mathbf{W}_2 \cdot_3 \mathbf{W}_3\|_2, \end{aligned} \tag{3}$$

wherein $\mathbf{W}_n \coloneqq \mathbf{U}_n\mathbf{U}_n^{\mathrm{T}}$. The tensors on the right-hand side of (3) are exactly those encountered in the successive sequential truncation steps and their spectral norms are bounded by the spectral norms of their unfoldings [14]. Unfolding these three tensors in modes one, two, and three, respectively, yields:

$$\begin{aligned}\breve{\epsilon}_2 &\leq \left\|\mathbf{T}_{(1)} - \mathbf{W}_1\mathbf{T}_{(1)}\right\|_2 + \left\|\left(\mathbf{T}_{(2)} - \mathbf{W}_2\mathbf{T}_{(2)}\right)(\mathbf{I}_{I_3} \otimes \mathbf{W}_1)\right\|_2 \\ &\quad + \left\|\left(\mathbf{T}_{(3)} - \mathbf{W}_3\mathbf{T}_{(3)}\right)(\mathbf{W}_2 \otimes \mathbf{W}_1)\right\|_2 \\ &= \|\breve{\mathbf{G}}_1\|_2 + \|\breve{\mathbf{G}}_2\|_2 + \|\breve{\mathbf{G}}_3\|_2.\end{aligned}$$

3.3. Choosing thresholds

RRURVDs estimate the rank in function of a threshold γ_n. Depending on the expected SNR of the tensor, γ_n has to be chosen accordingly. Using a result from matrix subspace tracking [25,28], we set γ_n equal to the expected size of the largest singular value of a matrix with i.i.d. zero-mean Gaussian distributed entries, multiplied by a factor $\alpha > 1$ to avoid false positives. In our case,

this leads to $\gamma_n = \alpha\sigma(\sqrt{m} + \sqrt{n})$ [20], wherein σ is an estimate of the standard deviation of the noise and m and n are the dimensions of the mode-n unfolding of $\mathcal{T}$. Thus, for mode 1, $m = I_1$ and $n = I_2 I_3$.

3.4. Computational cost

For both the TMLSVD and the TMLUTVD of a third-order tensor, two main algorithmic steps are required: first, the computation of three SVDs or three UTV-type decompositions and, second, the formation of the core tensor. The first step, which yields the factor matrices, is typically the most costly [27]. The second step consists of the multiplication of the given tensor with the transposes of these factor matrices and is identical for the TMLSVD and the TMLUTVD. Consequently, the difference in computational cost between the TMLSVD and the TMLUTVD boils down to differences in cost between a matrix SVD and UTV decomposition.

In practice, an SVD is typically computed through a reduction to bidiagonal form, which is a non-iterative phase, followed by a QR iteration, which is iterative. For an $(m \times n)$-matrix, with $m \geq n$, the computation of the SVD requires $\mathcal{O}(mn^2)$ flop [11]. UTV algorithms come in several different forms. Some, such as the RRQR and RRUTV algorithms start with a QR decomposition, which is a non-iterative procedure, and then apply iterative deflation steps to make the decomposition rank-revealing. Other forms of UTV algorithms, such as the HURV and SURV algorithms, directly obtain a rank-revealing UTV decomposition in a non-iterative way and can thus be computed in a finite number of steps. Rank-revealing UTV algorithms have the same asymptotic complexity ($\mathcal{O}(mn^2)$) as the SVD algorithm, but the coefficients with which the (mn^2)-term is scaled in their expressions are lower than for the SVD[2] [7, 11, 24, 28]. Intuitively, this makes sense, as an UTV decomposition imposes less structure on its factors than an SVD and therefore the computational cost of an UTV decomposition can be lower than that of the SVD. In addition, UTV algorithms can make better use of fast BLAS routines than the SVD, making them comparatively even faster in practice [18].

Considering $I = I_1 = I_2 = I_3$ and $R = R_1 = R_2 = R_3$, the computation of the MLSVD or MLUTVD of a tensor requires the computation of SVDs or UTV decompositions, respectively, of three tensor unfoldings. By substituting $m = I^2$ and $n = I$ in the expressions for the computational complexities of the SVD and the UTV decomposition, one obtains a complexity of $\mathcal{O}(I^4)$ for both. As only the principal subspace of dimension R is needed in the computation

[2] For both the SVD and the UTV algorithms, the precise coefficients depend on which matrices are explicitly computed.

of a TMLSVD or a TMLUTVD, the complexity can be reduced to $\mathcal{O}(RI^3)$ flop for both, with the coefficients for the UTV lower than those for the SVD.

Aside from the classical deterministic algorithms to compute matrix decompositions, approaches based on randomized numerical linear algebra [18] have been derived as well. These probabilistic methods iteratively sample and process parts of the data and can thus be applied to trade off accuracy for speed: the more data structure is exploited, the higher the computational cost, but also the higher the accuracy of the resulting estimated decomposition. Randomized algorithms exist for the SVD, see, e.g., [13], and the MLSVD [1, 3, 19]; randomized matrix UTV algorithms have been developed as well [16, 17]. We focus on deterministic methods below, as randomized methods lead to the same conclusion: like deterministic UTV methods are less computationally costly than deterministic SVD methods, the same is true for their stochastic counterparts [18].

4. Experiments

The experiments are performed on a computer with an Intel Core i7-6820HQ CPU at 2.70GHz and 16GB of RAM using MATLAB R2016b and Tensorlab 3.0. When computing TMLUTVDs in the experiments below, we use the rank-revealing URV decomposition [7] to obtain the UTV decompositions for each mode, but the results are very similar for other UTV algorithms.

In a first experiment, we compare the accuracy of the TMLURVD and the TMLSVD of a tensor $\mathcal{T} \in \mathbb{R}^{I \times I \times I}$, with $I \in \{20, 120\}$, having low multilinear rank. We construct a tensor $\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$ with multilinear rank $(3, 4, 5)$ by generating factor matrices $\mathbf{U}_1 \in \mathbb{R}^{I \times 3}$, $\mathbf{U}_2 \in \mathbb{R}^{I \times 4}$ and $\mathbf{U}_3 \in \mathbb{R}^{I \times 5}$ and a core $\mathcal{S} \in \mathbb{R}^{3 \times 4 \times 5}$ with entries sampled from the uniform distribution on $[0, 1]$. Gaussian noise is added with SNRs between 0 dB and 60 dB and 50 different tensors are considered. In Figure 4, the median angle (in radians) is shown between the subspaces generated by the true factor matrices $\mathbf{U}_n$ and by their corresponding estimated factor matrices $\hat{\mathbf{U}}_n$ for the worst mode. We also show the Frobenius and spectral errors between the perturbed tensor and the LMLRA $\hat{\mathcal{S}} \cdot_1 \hat{\mathbf{U}}_1 \cdot_2 \hat{\mathbf{U}}_2 \cdot_3 \hat{\mathbf{U}}_3$, along with the bounds from Section 3.1. Both bounds are good predictors for the accuracy of the LMLRA, but the Frobenius norm bound is tighter. From Figure 4, it can be seen that the TMLURVD achieves the same accuracy as the TMLSVD.

In a second experiment, we estimate the DOAs of sources impinging from the far-field on a uniform rectangular array (URA), using a data model similar to [12]. We consider R undamped exponentials impinging on a URA of dimensions $K \times L$ for N time instances. Because of the

URA structure, the entries of the third-order observation tensor $\mathcal{Y} \in \mathbb{C}^{K \times L \times N}$ satisfy

$$y_{kln} = \sum_{r=1}^{R} s_r(n) \mathrm{e}^{j(k-1)\mu_{r1}} \mathrm{e}^{j(l-1)\mu_{r2}} + n_{kln},$$

with j the imaginary unit, n_{kln} noise, $s_r(n)$ the complex amplitude of the rth exponential at the nth time instance and μ_r the coordinates of the rth spatial frequency vector. The latter depend on the DOAs of the R sources:

$$\mu_{r1} = 2\pi\delta_x \sin(\beta_r)\cos(\alpha_r)/\lambda$$
$$\mu_{r2} = 2\pi\delta_y \sin(\beta_r)\sin(\alpha_r)/\lambda,$$

with α_r and β_r the azimuths and elevations of the sources, δ_x and δ_y the inter-sensor spacing and λ the wavelength. This means that we can model $\mathcal{Y}$ as $\mathcal{Y} = \mathcal{A} \cdot_3 \mathbf{S}^{\mathrm{T}} + \mathcal{N}$, wherein $\mathbf{S} \in \mathbb{C}^{R \times N}$ holds the sources, $\mathcal{N}$ holds noise and $\mathcal{A} \in \mathbb{C}^{K \times L \times R}$ is a low-rank steering tensor with

$$a_{klr} = \mathrm{e}^{j(k-1)\mu_{r1}} \mathrm{e}^{j(l-1)\mu_{r2}}.$$

In the absence of noise, $\mathcal{Y}$ has rank R and no mode-n rank of $\mathcal{Y}$ exceeds R, see [4] for more details. To recover the source signals in the noisy case, we compute a multilinear rank-(R, R, R) approximation through the computation of a TMLSVD/TMLURVD $\mathcal{Y} \approx \mathcal{C} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \cdot_3 \mathbf{U}_3$. We compare the R-dimensional principal subspace of the column space of the resulting unfolding $\hat{\mathbf{A}}_{(3)}^{\mathrm{T}}$ of $\hat{\mathcal{A}} = \mathcal{C} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2 \in \mathbb{C}^{K \times L \times R_3}$ with the column space of the unfolding $\mathbf{A}_{(3)}^{\mathrm{T}}$ of the true steering tensor. This comparison is performed by looking at the principal angle between these two subspaces (using the MATLAB routine `subspace`).

For this experiment, we consider a URA with grid dimensions $K = L = 40$, $N = 100$ time steps and we generate $R = 6$ sources with azimuth and elevation angles α_r and β_r equal to the values in $[\,45\ -45\ 35\ 75\ 65\ -25\,]$ and $[\,45\ -45\ 15\ 85\ -35\ -40\,]$, respectively (in degrees). The real and imaginary parts of the source signals $s_r(n)$ are sampled from the standard normal distribution and we construct the observation tensor by forming the tensor $\mathcal{Y}$. We add Gaussian noise with an SNR between -10 and $40\,\mathrm{dB}$ and we then extract $\hat{\mathcal{A}}$ from a TMLSVD and a TMLURVD, using the thresholds as in Section 3.3 and $\alpha = 1.24$. In Figure 5, we show results for both the TMLSVD and the TMLURVD. It can be seen that both approaches result in the same median subspace errors over twenty trials.

Finally, in a third experiment, we compare the TMLSVD and the TMLURVD on a full-rank tensor. In this case, the tensor has no underlying low-rank structure, but the TMLSVD and TMLURVD can still be used to find an approximation of the tensor. The thresholds are

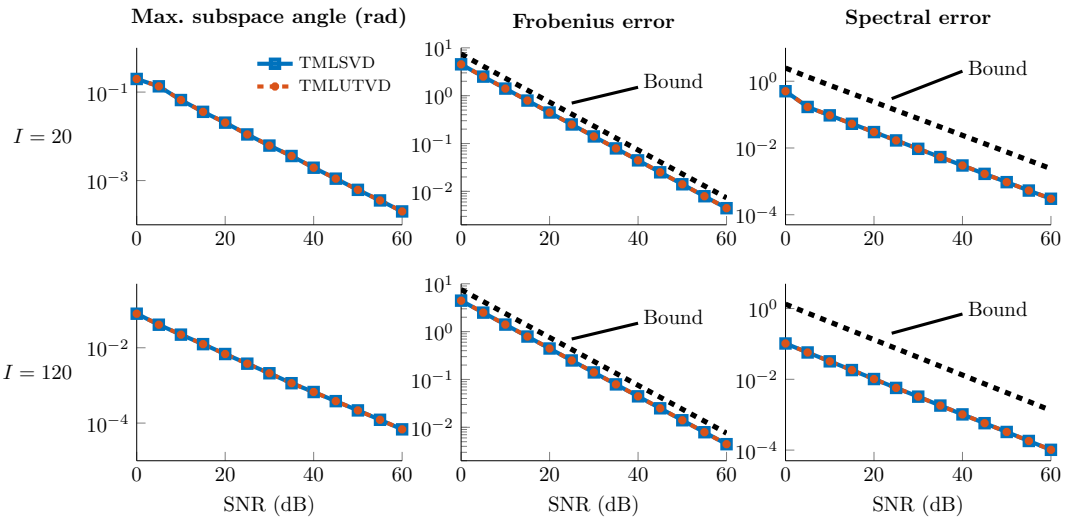

Figure 4: The TMLURVD and TMLSVD offer LMLRAs with similar accuracy. Medians (50 trials) of the maximal subspace angle (left) between the true ($\mathbf{U}_n$) and estimated ($\hat{\mathbf{U}}_n$) factors, and the Frobenius (center) and spectral (right) error between the perturbed tensor and the LMLRA $\hat{\mathcal{S}} \cdot_1 \hat{\mathbf{U}}_1 \cdot_2 \hat{\mathbf{U}}_2 \cdot_3 \hat{\mathbf{U}}_3$ for tensors of dimensions $I \times I \times I$ with $I = 20$ (top) and $I = 120$ (bottom). The bounds from Section 3.1 are also displayed.

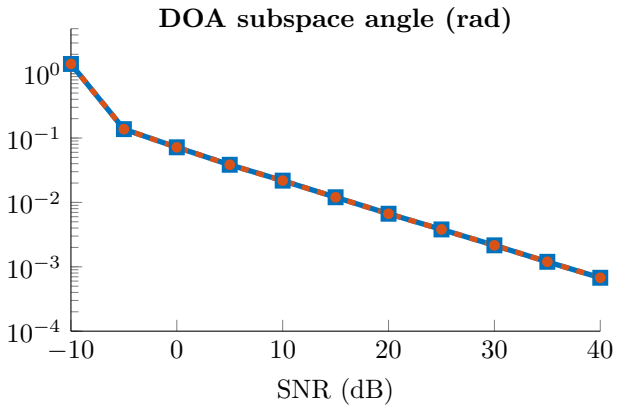

Figure 5: The median (20 trials) angle between the TMLSVD/TMLURVD subspaces and the true subspace for the DOA experiment.

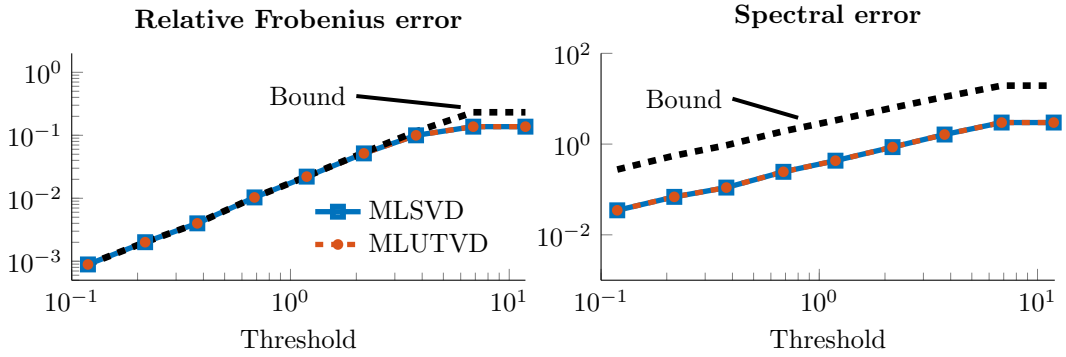

Figure 6: The TMLURVD and TMLSVD approximate full-rank tensors with similar accuracy. Medians (50 trials) of the relative Frobenius error (left) and spectral error (right) between the perturbed tensor and the LMLRA $\hat{\mathcal{S}} \cdot_1 \hat{\mathbf{U}}_1 \cdot_2 \hat{\mathbf{U}}_2 \cdot_3 \hat{\mathbf{U}}_3$ for full-rank tensors of dimensions $50 \times 50 \times 50$ with different threshold values. The bounds from Section 3.1 predict the errors well.

now used to specify the accuracy of the resulting approximation: lower thresholds yield better approximations. For each mode, we generate an orthogonal matrix $\mathbf{U}_n$ of dimensions 50×50 and multiply the rth column of $\mathbf{U}_n$ by $(51-r)/50$ to set the singular values of $\mathbf{U}_n$ to $(51-r)/50$, for $r = 1, 2, \ldots, 50$. We then form the tensor $\mathcal{T}$ of dimensions $50 \times 50 \times 50$ by using the $\mathbf{U}_n$ as factor matrices and generating a core tensor with entries sampled from the uniform distribution on $[0, 1]$. In Figure 6, the approximation errors in the Frobenius and spectral norms are shown for both the TMLSVD and the TMLURVD of T, computed for different values of the threshold $\gamma = \gamma_1 = \gamma_2 = \gamma_3$. As in the previous experiments, there is no noticeable difference in accuracy between the TMLSVD and the TMLURVD. However, here our main interest lies in the bounds from Section 3, which are displayed as well; they provide good estimates for the accuracy of the approximation. If the threshold was decreased even more, eventually the full mode-n subspaces would be included, meaning that the TMLSVD and TMLURVD would have full multilinear rank and thus represent the tensor exactly.

5. Conclusion

In this paper, we introduced a new tool for tensor compression, the TMLUTVD. Its computation is less costly than that of a TMLSVD, yet the TMLUTVD still provides the requested multilinear information in many practical applications. We derived bounds for the approximation error of the resulting LMLRA in both the Frobenius and spectral norms and illustrated that the accuracy of the TMLUTVD is comparable to that of the TMLSVD.

References

[1] S. Ahmadi-Asl, S. Abukhovich, M. G. Asante-Mensah, A. Cichocki, A.-H. Phan, T. Tanaka, and I. Oseledets. Randomized algorithms for computation of Tucker decomposition and higher order SVD (HOSVD). *IEEE Access*, 9:28684–28706, 2021.

[2] T. F. Chan. Rank revealing QR factorizations. *Linear Algebra and its Applications*, 88-89:67–82, 1987.

[3] M. Che and Y. Wei. Randomized algorithms for the approximations of Tucker and the tensor train decompositions. *Advances in Computational Mathematics*, 45(1):395–428, 2019.

[4] Y. Cheng, F. Roemer, O. Khatib, and M. Haardt. Tensor subspace tracking via Kronecker structured projections (TeTraKron) for time-varying multidimensional harmonic retrieval. *EURASIP Journal on Advances in Signal Processing*, 2014:123:1–14, 11 2014.

[5] A. Cichocki, C. Mandic, A-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer. Tensor decompositions for signal processing applications. From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32:145–163, 2015.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[7] R. D. Fierro and P. C. Hansen. Low-rank revealing UTV decompositions. *Numerical Algorithms*, 15(1):37–55, 1997.

[8] S. Friedland and L.-H. Lim. Nuclear norm of higher-order tensors. *Mathematics of Computation*, 87(311):1255–1281, 2018.

[9] D. Gerard and P. Hoff. A higher-order LQ decomposition for separable covariance models. *Linear Algebra and its Applications*, 505:57–84, 2016.

[10] G. Golub. Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3):206–216, 1965.

[11] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition, 2013.

[12] M. Haardt, F. Roemer, and G. Del Galdo. Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. *Signal Processing, IEEE Transactions on*, 56:3198 – 3213, 08 2008.

[13] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[14] S. He, Z. Li, and S. Zhang. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Mathematical Programming*, 125(2):353–383, 2010.

[15] F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematical Physics*, 6(1–4):164–189, 1927.

[16] M. F. Kaloorazi and R. C. de Lamare. Compressed randomized UTV decompositions for low-rank matrix approximations. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1155–1169, 2018.

[17] P.-G. Martinsson, G. Quintana-Ortí, and N. Heavner. RandUTV: A blocked randomized algorithm for computing a rank-revealing UTV factorization. *ACM Transactions on Mathematical Software*, 45(1):1–26, 03 2019.

[18] P.-G. Martinsson and J. A. Tropp. Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 29:403–572, 2020.

[19] R. Minster, A. K. Saibaba, and M. E. Kilmer. Randomized algorithms for low-rank tensor decompositions in the Tucker format. *SIAM Journal on Mathematics of Data Science*, 2(1):189–215, 2020.

[20] M Rudelson and R Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602, Hyderabad, India, August 2010. World Scientific.

[21] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

[22] G. W. Stewart. An updating algorithm for subspace tracking. *IEEE Transactions on Signal Processing*, 40:1535–1541, 1992.

[23] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[24] A.-J. van der Veen. Subspace tracking using a constrained hyperbolic URV decomposition. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP '98)*, volume 4, pages 1953–1956, 1998.

[25] M Vandecappelle and L De Lathauwer. Low multilinear rank updating. In *Proceedings. of the 53rd Asilomar Conference on Signals, Systems and Computers (ASILOMAR 2019)*, pages 437–441, Pacific Grove, CA, USA, Nov 2019.

[26] N. Vannieuwenhoven, K. Meerbergen, and R. Vandebril. Computing the gradient in optimization algorithms for the CP decomposition in constant memory through tensor blocking. *SIAM Journal on Scientific Computing*, 37(3):C415–C438, 2015.

[27] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen. A new truncation strategy for the higher-order singular value decomposition. *SIAM Journal on Scientific Computing*, 34(2):A1027–A1052, 2012.

[28] M. Zhou and A. van der Veen. Stable subspace tracking algorithm based on a Signed URV decomposition. *IEEE Transactions on Signal Processing*, 60(6):3036–3051, June 2012.